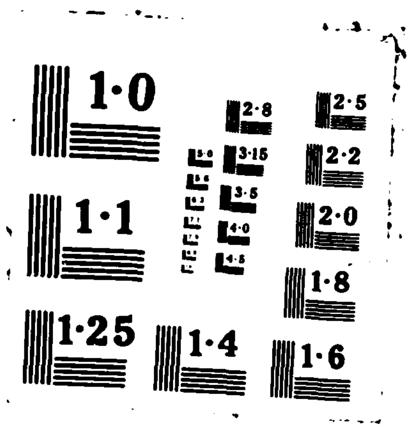
AD-A188 960 ESTIMATION OF THE PROBABILITY OF A LONG TIME TO THE FIRST ENTRANCE TO A STATE IN A SEMI-MARKOV MODEL(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA P A JACOBS 1/1

UNCLASSIFIED SEP 87 NPS55-87-012 F/G 12/3 NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
4·0
2·0
4·5
1·8
1·25
1·4
1·6

NPS55-87-012

NAVAL POSTGRADUATE SCHOOL

Monterey, California

ESTIMATION OF THE PROBABILITY OF A LONG TIME TO THE FIRST ENTRANCE TO A STATE IN A SEMI-MARKOV MODEL

PATRICIA A. JACOBS

SEPTEMBER 1987

Approved for public release; distribution unlimited.

Prepared for:
Office of Naval Research
Arlington, VA 22217

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPS55-87-012	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL (If applicable) Code 55	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable) Code 430	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N0001487WR24005
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR014-05-01	4114531-02	

11. TITLE (Include Security Classification)
ESTIMATION OF THE PROBABILITY OF A LONG TIME TO THE FIRST ENTRANCE TO A STATE IN A SEMI-MARKOV MODEL

12. PERSONAL AUTHOR(S)
JACOBS, PATRICIA A.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	Sep 1987	27

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Estimation of the survivor function of a first passage time; semi-Markov process; empirical moment generating function; asymptotic exponentiality; strong law; central limit theorem;

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The problem of estimating the log-survivor function for a fixed time t of a first passage time for a semi-Markov process is considered. A nonparametric estimator based on an exponential approximation to the survivor function is introduced and its asymptotic properties are studied. Small sample behavior of this estimator is studied using simulation

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL P. A. Jacobs	22b. TELEPHONE (Include Area Code) (408)646-2258 22c. OFFICE SYMBOL 55Jc

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

U.S. Government Printing Office: 1986-606-243

18. bootstrap; jackknife

1. INTRODUCTION

Finite state space semi-Markov processes find application in many areas including queueing, reliability, and clinical trials; (cf Cinlar [1975], Weiss and Zelen [1965], Cox [1985]). Often interest in the application centers on the distribution of a first-passage time to a state or a set of states representing for example the lifetime of a system or the end of a busy period of a server.

In this paper we consider the problem of the estimation of the log survivor function of a first passage time for a fixed time t. We suppose that observational data are known about the semi Markov process in question. Although the approach and results are given concretely for one particular semi-Markov process, they apply more widely.

To be specific, consider N individuals. Let $X_t(i)$ be the state of ith individual at time t. We will assume $\{X_t(i);t\geq 0\}$ are independent identically distributed processes having the same probability law as $\{X_t;t\geq 0\}$. The process $\{X_t;t\geq 0\}$ is a semi-Markov process with three states $\{0,1,2\}$. State 0 is absorbing and $X_0=1$. The sojourn time in state i has a distribution function F_i. Upon leaving state 1, the process transitions to state 0 with probability θ and to state 2 with probability $1-\theta$. From state 2 the process transitions to state 1 with probability 1.

Let

$$D = \inf\{t\geq 0: \ X_t = 0\}$$

the entrance time to state 0.

The problem is to estimate the logarithm of $P\{D>t\}$, for fixed t, from data obtained by observing the N individuals. Three possible estimation procedures will be considered. The three estimators use different amounts of information concerning the process. One procedure uses only the observed entrance time to state 0 for the N individuals to estimate the empirical distribution function of D. Another procedure makes parametric assumptions concerning the sojourn time distributions F_i and uses maximum likelihood to estimate the probability. A third approach uses an exponential approximation to $P\{D>t\}$ and empirical distributions to estimate F_i.

In Section 2, the estimation procedure based on an exponential approximation to $P\{D>t\}$ is described and asymptotic results are obtained for it.

In Section 3 the other two estimation procedures are described and the results of a simulation experiment are given. The experiment was conducted to compare the performance of confidence intervals for the three estimators for small and moderate numbers of individuals. It is found that the procedure based on the maximum likelihood is the most efficient if the underlying model is correct. However, maximum likelihood confidence interval coverage is sensitive to incorrect model assumptions. The nonparametric confidence intervals based on the exponential approximation have the correct coverage as long as the time t is not too small. The conservative binomial confidence intervals based only on the observed times of absorption are the least efficient and tend to overcover.

2. AN ASYMPTOTIC RENEWAL ESTIMATE

The probability $P\{D>t\}$ for the model described in the Introduction

satisfies the equation

$$P\{D>t\} = g(t) + (1-\theta) \int_0^t (F_1 * F_2)(ds)\, P\{D>t-s\} \qquad (2.1)$$

with

$$g(t) = \bar{F}_1(t) + (1-\theta) \int_0^t F_1(ds)\, \bar{F}_2(t-s). \qquad (2.2)$$

This is a renewal equation with defective inter-renewal distribution

$$L(t) = (1-\theta)(F_1 * F_2)(t). \qquad (2.3)$$

Following Feller [1966] page 326, assume there exists κ which satisfies the following equation

$$\int_0^\infty e^{\kappa t}\, L(dt) = (1-\theta)\, \phi_1(\kappa)\, \phi_2(\kappa) = 1 \qquad (2.4)$$

where

$$\phi_i(\xi) = \int_0^\infty e^{\xi t}\, F_i(dt). \qquad (2.5)$$

If $(F_1 * F_2)$ is not arithmetic, then under certain integrability conditions

$$\lim_{t\to\infty} e^{\kappa t}\, P\{D>t\} = \frac{b}{\mu} \qquad (2.6)$$

where

$$\mu = \int_0^{\infty} s e^{\kappa s} L(ds) \tag{2.7}$$

and

$$b = \int_0^{\infty} e^{\kappa s} g(s)ds = \frac{\theta}{\kappa} \phi_1(\kappa). \tag{2.8}$$

Following an approach that was found useful in another context, (Gaver and Jacobs ([1986]), an estimator for $P(D>t)$ can be obtained by using estimates of κ, b and μ in (2.6). Specifically, let $\{S_n(i), n=1, \ldots, M_i\}$ be the collection of sojourn times in state i for the N individuals and let R be the number of transitions from state 1 to state 2 for all the N individuals.

Put $\hat{\theta} = \frac{N}{N+R}$ and

$$\hat{\phi}_i(\xi) = \frac{1}{M_i} \sum_{k=1}^{M_i} e^{\xi S_k(i)}, \tag{2.9}$$

i = 1,2.

An estimator of κ is the solution, $\hat{\kappa}$, to the equation

$$(1-\hat{\theta})\, \hat{\phi}_1(\xi)\hat{\phi}_2(\xi) = 1. \tag{2.10}$$

Notice that the left hand side of equation (3.10) is nondecreasing in ξ and at $\xi = 0$ equals $(1-\hat{\theta}) \leq 1$. Hence (2.10) has a unique solution which can be found numerically.

The estimators for the other quantities in (2.6) are

$$\hat{\mu} = (1-\hat{\theta})\frac{1}{M_1M_2}\sum_{k=1}^{M_1}\sum_{j=1}^{M_2} \exp\{\kappa(S_k(1)+S_j(2))\}\ (S_k(1) + S_j(2)) \quad (2.11)$$

$$\hat{b} = \frac{\hat{\theta}}{\kappa}\ \hat{\Phi}_1(\hat{\kappa}). \quad (2.12)$$

The asymptotic exponential estimator for $P\{D>t\}$ is

$$\hat{P}_A\{D>t\} = \frac{\hat{b}}{\mu} e^{-\hat{\kappa}t}. \quad (2.13)$$

2.1 Large Sample Properties of $\hat{\kappa}$ and $\ell n\ \hat{P}_A\{D>t\}$

Let

$$\hat{f}(\xi,\alpha) = 1 - (1-\alpha)\hat{\Phi}_1(\xi)\hat{\Phi}_2(\xi) \equiv 1 - (1-\alpha)\hat{h}(\xi).$$

Expand $\hat{h}(\hat{\kappa})$ in Taylor series about the solution κ of (2.4). Since $\hat{f}(\hat{\kappa},\hat{\theta}) = 0$,

$$0 = \hat{f}(\hat{\kappa},\hat{\theta}) = 1 - (1-\hat{\theta})\ [\hat{h}(\kappa) + (\hat{\kappa}-\kappa)\hat{h}'(\kappa) + \tfrac{1}{2}(\hat{\kappa}-\kappa)^2\ \hat{h}''(a\kappa)] \quad (2.14)$$

$$= \hat{f}(\kappa,\theta) + (\hat{\theta}-\theta)\hat{h}(\kappa) - (1-\hat{\theta})(\hat{\kappa}-\kappa)[\hat{h}'(\kappa) + \tfrac{1}{2}(\hat{\kappa}-\kappa)\hat{h}''(a\kappa)].$$

Thus

$$\hat{\kappa}-\kappa = \frac{\hat{f}(\kappa,\theta) + (\hat{\theta}-\theta)\ \hat{h}(\kappa)}{B} \quad (2.15)$$

where

$$B = (1-\hat{\theta})[\hat{h}'(\kappa) + \frac{1}{2}(\hat{\kappa}-\kappa)\ \hat{h}''(a\kappa)], \tag{2.16}$$

and a is a constant and θ is the probability of a transition from state 1 to state 0. By the strong law of large numbers, as $N\to\infty$ $\hat{f}(\kappa,\theta)\to 0$ and $\hat{\theta}-\theta\to 0$. Further, as $N\to\infty$, with probability one

$$\hat{h}'(\kappa) \to (1-\theta) \int_0^\infty se^{\kappa s}(F_1 * F_2)(ds) \tag{2.17}$$

and

$$\hat{h}(\kappa) \to \Phi_1(\kappa)\ \Phi_2(\kappa). \tag{2.18}$$

Therefore, $\hat{\kappa}-\kappa\to 0$ as $N\to\infty$, with probability 1 so the estimator $\hat{\kappa}$ is consistent.

To obtain a central limit theorem for $\hat{\kappa}$, the following moment condition is needed. Assume

$$\int_0^\infty e^{2\kappa s}\ F_1 * F_2(ds) < \infty. \tag{2.19}$$

Since $f(\kappa,\theta) = 0$,

$$\sqrt{N}\ [\ \hat{f}(\kappa,\theta)] = \sqrt{N}\ [\ \hat{f}(\kappa,\theta) - f(\kappa,\theta)]$$

$$= \sqrt{N}\ (1-\theta)\ [\hat{\Phi}_1(\kappa)\hat{\Phi}_2(\kappa) - \Phi_1(\kappa)\Phi_2(\kappa)]$$

$$= \sqrt{N}\ (1-\theta)\ \{[\hat{\Phi}_1(\kappa) - \Phi_1(\kappa)]\ \hat{\Phi}_2(\kappa) + \Phi_1(\kappa)[\hat{\Phi}_2(\kappa)-\Phi_2(\kappa)]\}. \tag{2.20}$$

Thus, (2.19) and the central limit theorem imply that $\sqrt{N}\ [\hat{f}(\kappa,\theta)]$ is asymptotically normal as $N\to\infty$ with mean 0 and variance

$$\mathrm{Var}_f = (1-\theta)\ \theta\{(1-\theta)\Phi_2(\kappa)^2\ \mathrm{Var}[e^{\kappa T_1}] + \Phi_1(\kappa)^2\ \mathrm{Var}[e^{\kappa T_2}]\} \tag{2.21}$$

where T_i is a random variable with distribution F_i.

Further the central limit theorem implies that as $N\to\infty$, $\sqrt{N}\ [\hat{\theta}-\theta]$ is asymptotically normal with mean 0 and variance $V_b = \theta^2(1-\theta)^2 + \theta^3(1-\theta)$.

Hence, as $N\to\infty$, $\sqrt{N}\ (\hat{\kappa}-\kappa)$ is asymptotically normal with mean 0 and variance

$$V_\kappa = (V_f + (\Phi_1(\kappa)\ \Phi_2(\kappa))^2\ V_b)[(1-\theta)\ E[(T_1+T_2)e^{\kappa(T_1+T_2)}]]^{-2} \tag{2.22}$$

A similar argument shows that $\ell n\ \hat{P}_A\{D>t\}$ is also asymptotically normally distributed as $N\to\infty$.

3. SIMULATION RESULTS

A simulation experiment was carried out to investigate small and moderate sample size behavior of confidence interval procedures for three estimators of $\ell n\ P\{D>t\}$. The following three estimators use different amounts of information about the semi-Markov processes $\{X_t(i);\ t\geq 0\}$, $i=1,\ldots,N$, for the N individuals

3.1 The Binomial Estimator

An estimator which uses only the observed entrance times into state 0 for the N individuals, $d_1, d_2, \ldots, d_N$, is

$$\hat{P}_B\{D>t\} = \frac{1}{N}\sum_{i=1}^{N} 1_{(t,\infty)}(d_i) \qquad (3.1)$$

where

$$1_{(t,\infty)}(d_i) = \begin{cases} 1 & \text{if } d_i>t, \\ 0 & \text{otherwise.} \end{cases}$$

The distribution of the estimator $\hat{P}_B\{D>t\}$ is binomial and a binomial confidence interval $[L_B,U_B]$ can be constructed for $P\{D>t\}$ cf. Larson [1982] p 397. The binomial confidence interval for $\ell n\ P\{D>t\}$ is $[\ell n\ L_B, \ell n\ U_B]$.

3.2 The Asymptotic Exponential Estimator.

The asymptotic exponential estimator for $P\{D>t\}$ is given by $\hat{P}_A\{D>t\}$ of (2.13). The estimator for the $\ell n\ P\{D>t\}$ is $\ell n\ \hat{P}_A\{D>t\}$ and is asymptotically normal as $N\to\infty$. However, since the simulation experiment is for small to moderate numbers of individuals, two nonparametric methods are used to obtain confidence intervals.

The *jackknife* is a procedure originally introduced by Quenouille [1956] for bias reduction, and adapted by Tukey [1958] to obtain approximate confidence intervals. Suppose interest is on a parameter β that is estimated by $\hat{\beta}$ using a complex calculation from data. The idea is that of assessing variability by recomputing $\hat{\beta}$ after removing independent subgroups of data and then using the recomputed $\hat{\beta}$ values to estimate a variance which is in turn applied to state a two-sided confidence interval that contains the true β with

specified confidence; see Efron [1982] and his more recent work, or Mosteller and Tukey [1977] for more details. For simulations involving N=20 individuals, the jth subgroup consists of all data corresponding to the jth individual. For simulations involving N=50 individuals the first subgroup consists of all data corresponding to the first five individuals the second subgroup of all data corresponding to the second five individuals, etc. Some simulations for N=50 individuals were run with the jth subgroup consisting of the data for the jth individual alone; according to some work these should be an improvement on those with 5 in a subgroup. However, the resulting confidence intervals differed little from those obtained by leaving out 5 individuals at a time.

The bootstrap is an alternate method that may be used to assess uncertainty of estimators; cf. Efron [1982] and his more recent work. In the simulations reported here a bootstrap replication is generated as follows. Let R_n be the number of transitions from state 1 to state 2 for individual n. Let $\{S_k(i)\}$ be the collection of all sojourn times in state i for all individuals. To generate bootstrap data for one individual, one observation is drawn at random with replacement from $\{R_n\}$; call the observation r; r+1 observations are drawn at random with replacement from the collection of state-1 sojourn times $\{S_k(1)\}$; r observations are drawn at random with replacement from the collection of state-2 sojourn times $\{S_k(2)\}$. Bootstrap data is generated for N individuals and the estimate $\ell n\ P_A\{D>t\}$ is computed using this data. This completes one bootstrap replication. B = 100 bootstrap replications are done. The B = 100 bootstrap estimates of $\ell n\ \hat{P}_A\{D>t\}$ are ordered. A 100 x $(1-\alpha)$ % confidence interval is constructed using the $\alpha/2$ and $1-\alpha/2$ quantiles of the bootstrap estimates.

3.3 The Maximum Likelihood Estimator

In what follows, we assume the sojourn time in state i has an exponential distribution with mean $\frac{1}{\rho_i}$. The log-likelihood function under these assumptions is

$$L = \sum_{n=1}^{N} [R_n \ell n\,(1-\theta) + \ell n\,\theta + R_n \ell n\,\rho_2 + (1+R_n)\ell n\,\rho_1 - \rho_1 T_n(1) - \rho_2 T_n(2)] \tag{3.2}$$

where R_n is the number of transitions from state 1 to state 2 for individual n and $T_n(i)$ is the total time spent in state i by individual n. The maximum likelihood estimators are

$$\hat{\theta} = \frac{N}{N+R} \tag{3.3}$$

$$\hat{\rho}_1 = \frac{N+R}{T(1)} \tag{3.4}$$

$$\hat{\rho}_2 = \frac{R}{T(2)} \tag{3.5}$$

where $R = \sum_{n=1}^{N} R_n$, $T(1) = \sum_{n=1}^{N} T_n(1)$, and $T(2) = \sum_{n=1}^{N} T_n(2)$.

Let D be the entrance time to state 0 for an individual; then

$$S(t) \equiv P(D>t) = \{\frac{\lambda_2+\rho_2}{\lambda_2} \exp\{\lambda_2 t\} - \frac{\lambda_1+\rho_2}{\lambda_1} \exp\{\lambda_1 t\}\} \frac{\theta\rho_1}{\lambda_1-\lambda_2} \tag{3.6}$$

where λ_1, λ_2 are the roots of the equation

$$\theta\rho_1\rho_2 + y(\rho_1+\rho_2) + y^2 = 0. \qquad (3.7)$$

The maximum likelihood estimator for $\ell n\ P\{D>t\}$ can be obtained from (3.6) and (3.7) by replacing ρ_1, ρ_2 and θ by their maximum likelihood estimators and taking the logarithm of the resulting estimator of $P\{D>t\}$. The distribution of the resulting estimator is asymptotically normal as $N\to\infty$. Since the maximum likelihood estimators $\hat{\theta}, \hat{\rho}_1$, and $\hat{\rho}_2$ are uncorrelated, the asymptotic variance is

$$\mathrm{Var}[\ell n\ \hat{P}\{D>t\}] = \frac{1}{S(t)^2}\,[\mathrm{Var}(\hat{\theta})\ (\frac{\partial S}{\partial \theta})^2 + \mathrm{Var}(\hat{\rho}_1)(\frac{\partial S}{\partial \rho_1})^2 + \mathrm{Var}(\hat{\rho}_2)(\frac{\partial S}{\partial \rho_2})^2\ ]. \qquad (3.8)$$

Confidence intervals are constructed using the asymptotic normality of the maximum likelihood estimator and the asymptotic variance with parameter values being replaced by their corresponding estimates. Confidence intervals based on the maximum likelihood ratio test may also be constructed [cf. Cox and Hinkley (1974) pp. 343]. While these intervals are asymptotically equivalent to the normal confidence intervals that are computed, they may be better for sample sizes.

3.4 The Simulation Experiment

All simulations are carried out on an IBM 3033AP computer at the Naval Postgraduate School using the LLRANDOMII random number generating package; [cf

Lewis and Uribe (1981)]. Some details of the simulation are given below. A more complete account can be found in Kim [1987]. In each replication estimates and confidence intervals are computed for the logarithm of P{D>t} using the procedures described in subsections 3.1-3.3. Each simulation is replicated 300 times. For each procedure, the number of confidence intervals covering the true value of ℓn P{D>t} is recorded. Also recorded are the number of intervals that are too low (true ℓn P{D>t} > U, the upper endpoint of the interval) and too high (true ℓn P{D>t} < L, the lower endpoint of the interval). The average length of the confidence intervals is also computed.

Tables 1-8 present results from simulations for N=20 and N=50 individuals. The first row in each table lists the confidence interval procedures: the binomial confidence interval for logarithm of estimate (3.1) (BIN); the maximum likelihood interval (MLE); the jackknife confidence interval for the asymptotic estimate, (JK); and the bootstrap confidence interval for the asymptotic estimate, (BT). Nominal 80% and 90% confidence interval for ℓn P{D>t} are computed using each procedure for each time t = 0.5, 1.0, 1.5, 2, 3, 4. Confidence limits that are greater than 0 are set equal to 0.

Tables 1-4 present results for the case in which the sojourn times in state 1 have an exponential distribution with ρ_1^{-1} = 1 and the sojourn times in state 2 have an exponential distribution with mean ρ_2^{-1} = 0.1. The values of P{D>t} for the model for t = 0.5, 1.0, 1.5, 2, 3, 4 are given in Appendix A.

Tables 1-2 show results for simulations with N=20 individuals. Table 1 shows coverage results. Under each procedure is given the number of intervals that cover the true value (C), the number of intervals that are too low (L) and the number of intervals that are too high (H). In parentheses

next to these numbers are given the corresponding fractions. If the true confidence level of a confidence interval procedure is 80%, (respectively 90%), then a 95% confidence interval for the fraction of the 300 replications that cover the true value is $0.80 \pm 1.96 \sqrt{\frac{1}{300}(.2)(.8)} = [.755, .845]$, (respectively $.9 \pm 1.96 \sqrt{\frac{1}{300}(.1)(.9)} = [.866, .934]$).

Table 2 shows the average length of confidence intervals for $\ell n\ P\{D>t\}$ for each t and each procedure. In parentheses under the average length is the estimated standard error of the average length.

The conservative binomial procedure tends to overcover and has the largest average confidence interval length. The maximum likelihood procedure has the correct coverage and smallest average length. The confidence intervals based on the asymptotic renewal estimate show the correct coverage for the times larger than t = 0.5 and 1.0. When the jackknife and bootstrap intervals do not cover they tend to be too high. The average length of the jackknife and bootstrap intervals is between those of the binomial and maximum likelihood intervals. The average length of the jackknife intervals is greater than those for the bootstrap, suggesting that the latter is preferred; however computational cost may be higher.

Tables 3 and 4 present coverage results and average confidence interval lengths for a simulation experiment with N=50 individuals. The coverage results are similar to those of Table 1. As anticipated, the average confidence interval lengths tend to be smaller for N = 50 individuals than for N=20 individuals. The average bootstrap interval length is closer to that of the MLE interval for times larger than 1 for N=50 individuals than for N=20 individuals.

Tables 5-8 report results of a simulation experiment in which the sojourn

times in states 1 and 2 have the following distributions. The sojourn time in state 2 has an exponential distribution with mean $\frac{1}{\rho_2} = 0.1$ as before. The sojourn time in state 1 is the sum of two independent exponential random variables each having mean $\frac{1}{2}$. Thus, the mean sojourn time in each state is the same as for the model used in Tables 1-4. Values of P{D>t} for various values of t for this model are given in Appendix B. Estimators and confidence intervals for ℓn P{D> } are computed as before. In particular the maximum likelihood estimators assume that the sojourn times in state 1 have an exponential distribution with mean 1. The number of individuals is N=20 for the simulation in Tables 5-6. There are N=50 individuals in the simulations of Tables 7-8. The coverage results contained in Tables 5 and 7 show that the maximum likelihood confidence intervals are sensitive to underlying model assumptions. Incorrect model assumptions lead to both overcoverage and undercoverage. The confidence intervals based on the exponential approximation have the correct coverage for the times larger than 0.5. The average length of the jackknife and bootstrap intervals is smaller than those for the binomial. The average length of the bootstrap interval is smaller than those for the jackknife.

4. CONCLUSIONS

The simulation results indicate that the maximum likelihood estimator is most efficient when the correct model is used. As has been noted elsewhere, the MLE estimator is rather sensitive to incorrect model assumptions. The simulation results for the asymptotic exponential estimator are somewhat surprising in that it appears that the time t does not have to be very large in order for the asymptotic exponential estimator to work well. Both the

nonparametric bootstrap and jackknife confidence intervals for the asymptotic exponential estimator have the correct coverage for moderately large t. As implemented the bootstrap intervals tend to have a smaller average length than the jackknife. However, neither a complete jackknife nor a bootstrap with more than 100 replications has been done due to the massive amount of computation involved. As expected, the binomial estimator produces conservative confidence intervals which tend to overcover. However, the binomial confidence intervals are appropriate for all times and all semi-Markov models.

5. ACKNOWLEDGEMENT

The author gratefully acknowledges the research support by the Probability and Statistics Program of the Mathematical Sciences Division of the Office of Naval Research. She also thanks D. P. Gaver and D. R. Cox for their comments.

REFERENCES

E. Cinlar, Introduction to Stochastic Processes, Prentice-Hall, Englewood Cliffs, N.J., 1975.

D. R. Cox, Some remarks on semi-Markov processes in medical statistics, International Symposium on Semi-Markov Processes and Their Applications, ed. J. Janssen, Plenum, New York, 1986.

D. R. Cox and D. V. Hinkley, Theoretical Statistics, Chapman and Hill, London, 1974.

B. Efron, The Jackknife, the Bootstrap, and other Resampling Plans, SIAM-NSF, 1982.

W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, John Wiley and Sons, 1966.

D. P. Gaver and P. A. Jacobs, Nonparametric estimation of the probability of a long delay in the M/G/1 queue. Naval Postgraduate School Technical Report, Monterey, CA 1986. To appear J. Royal Statist Soc Ser B.

S. W. Kim, A Simulation Study of Estimates of a first Passage Time Distribution for a semi-Markov process, M.S. Thesis, Naval Postgraduate School, March 1987.

H. J. Larson, Introduction to Probability and Statistical Inference, 3rd Edition, John Wiley and Sons, Inc., New York, 1982.

P. A. W. Lewis and L. Uribe, The new Naval Postgraduate School random number package LLRANDOM II, Naval Postgraduate School Technical Report, NPS55-81-005, Monterey, CA.

F. Mosteller and J. W. Tukey, Data Analysis and Regression, Addison-Wesley, 1977.

M. H. Quenouille, Notes on bias in estimation, Biometrika, 4, (1956), 353-360.

J. W. Tukey, Bias and confidence in not-quite large samples, (abstract), Ann Math Stat. 29, (1958), 614.

G. H. Weiss and M. Zelen, A semi Markov model for clinical trials, J. Appl. Prob. 2, (1965), 264-285.

Appendix A

Values of P{D>t} for the semi-Markov Model
with exponential sojourn time in state 1
$\rho_1=1$, $\rho_2=10$, $\theta=0.5$

time t	P{D>t}
0.5	.7866
1.0	.6203
1.5	.4891
2.0	.3857
3.0	.2399
4.0	.1492

Appendix B

Values of P{D>t} for the semi-Markov Model with nonexponential sojourn time in state 1

$\rho_1=2,\ \rho_2=10,\ \rho_3=2,\ \theta=0.5$

time t	P{D>t}
0.5	.8652
1.0	.6743
1.5	.5157
2.0	.3930
3.0	.2278
4.0	.1320

Table 1

Coverage Results for Confidence Intervals for $\ln P\{D>t\}$
Exponential Sojourn Times
$N=20,\ \theta=0.5,\ \rho_1=1.0,\ \rho_2=10.0$

Time	Confidence Level	Coverage	BIN	MLE	JK	BT
0.5	80%	H	14(.05)	24(.08)	106(.35)	140(.47)
		C	272(.91)	251(.84)	174(.58)	160(.53)
		L	14(.05)	25(.08)	20(.07)	0(.0)
	90%	H	4(.01)	14(.05)	78(.26)	99(.33)
		C	282(.94)	276(.92)	211(.70)	201(.67)
		L	14(.05)	10(.03)	11(.04)	0(.00)
1.0	80%	H	17(.06)	24(.08)	72(.24)	66(.22)
		C	250(.83)	250(.83)	210(.70)	224(.75)
		L	33(.11)	26(.09)	18(.06)	10(.03)
	90%	H	7(.02)	14(.05)	46(.15)	33(.11)
		C	275(.92)	277(.92)	245(.82)	254(.86)
		L	18(.06)	9(.03)	9(.03)	8(.03
1.5	80%	H	15(.05)	24(.08)	48(.16)	40(.13)
		C	265(.88)	250(.83)	247(.82)	236(.79)
		L	20(.07)	26(.09)	5(.02)	24(.08)
	90%	H	15(.05)	14(.05)	33(.11)	19(.06)
		C	278(.93)	277(.92)	266(.89)	266(.89)
		L	7(.02)	9(.03)	1(.00)	15(.05)
2.0	80%	H	11(.04)	24(.08)	50(.17)	27(.09)
		C	267(.89)	250(.83)	245(.83)	240(.80)
		L	22(.07)	26(.09)	5(.02)	33(.11)
	90%	H	11(.04)	14(.05)	28(.09)	10(.03)
		C	280(.93)	277(.92)	271(.90)	267(.89)
		L	9(.03)	9(.03)	1(.00)	23(.08)
3.0	80%	H	23(.08)	23(.08)	40(.13)	19(.06)
		C	271(.90)	251(.84)	252(.84)	238(.79)
		L	6(.02)	26(.09)	8(.03)	43(.14)
	90%	H	8(.03)	14(.05)	22(.07)	9(.03)
		C	286(.95)	277(.92)	276(.92)	260(.87)
		L	6(.02)	9(.03)	2(.01)	31(.10)
4.0	80%	H	19(.06)	23(.08)	36(.12)	15(.05)
		C	271(.90)	251(.84)	253(.84)	231(.77)
		L	10(.03)	26(.09)	11(.04)	54(.18)
	90%	H	4(.01)	14(.05)	21(.07)	7(.03)
		C	289(.96)	277(.92)	277(.92)	261(.87)
		L	7(.02)	9(.03)	2(.01)	32(.11)

Table 2

Average Length of Confidence Intervals for ℓn P{D>t}
Exponential Sojourn Times
$N=20$, $\theta=0.5$, $\rho_1=1.0$, $\rho_2=10.0$

Time	Confidence Level	BIN	MLE	JK	BT
0.5	80%	.370	.144	.208	.185
		(.005)	(.002)	(.005)	(.003)
	90%	.460	.186	.271	.245
		(.007)	(.002)	(.006)	(.005)
1.0	80%	.545	.288	.307	.290
		(.008)	(.004)	(.006)	(.005)
	90%	.680	.370	.400	.382
		(.009)	(.005)	(.008)	(.006)
1.5	80%	.716	.431	.467	.436
		(.010)	(.006)	(.011)	(.008)
	90%	.893	.553	.608	.579
		(.012)	(.007)	(.014)	(.010)
2.0	80%	.915	.574	.645	.597
		(.016)	(.008)	(.015)	(.011)
	90%	1.14	.737	.840	.793
		(.020)	(.010)	(.019)	(.015)
3.0	80%	1.38	.860	1.02	.932
		(.033)	(.011)	(.024)	(.018)
	90%	1.71	1.10	1.33	1.24
		(.038)	(0.15)	(:032)	(.025)
4.0	80%	2.11	1.15	1.40	1.27
		(.069)	(.015)	(.033)	(.026)
	90%	2.58	1.47	1.82	1.70
		(.076)	(.020)	(.044)	(.034)

Table 3

Coverage Results for Confidence Intervals for $\ell n\ P\{D>t\}$
Exponential Sojourn Times
$N=50,\ \theta=0.5,\ \rho_1=1.0,\ \rho_2=10.0$

Time	Confidence Level	Coverage	BIN	MLE	JK	BT
0.5	80%	H	17(.06)	27(.09)	119(.40)	151(.50)
		C	260(.87)	245(.83)	167(.56)	149(.50)
		L	23(.08)	28(.09)	14(.05)	0(.00)
	90%	H	5(.02)	14(.05)	83(.28)	110(.37)
		C	288(.96)	274(.91)	209(.70)	190(.63)
		L	7(.02)	12(.04)	8(.03)	0(.00)
1.0	80%	H	28(.09)	27(.09)	70(.23)	80(.27)
		C	245(.82)	245(.82)	216(.72)	213(.71)
		L	27(.09)	28(.09)	14(.05)	7(.02)
	90%	H	8(.03)	14(.05)	45(.15)	45(.15)
		C	282(.94)	274(.91)	274(.82)	250(.83)
		L	10(.03)	12(.04)	8(.03)	5(.02)
1.5	80%	H	19(.06)	27(.09)	53(.17)	46(.15)
		C	262(.87)	245(.82)	235(.78)	233(.78)
		L	19(.06)	28(.09)	12(.04)	21(.07)
	90%	H	11(.04)	14(.05)	32(.11)	23(.08)
		C	278(.93)	274(.91)	262(.87)	266(.89)
		L	11(.04)	12(.04)	6(.02)	11(.04)
2.0	80%	H	18(.06)	27(.09)	47(.16)	31(.10)
		C	260(.87)	245(.82)	240(.80)	243(.81)
		L	22(.07)	28(.09)	13(.04)	26(.09)
	90%	H	9(.03)	14(.05)	22(.07)	15(.05)
		C	277(.92)	274(.91)	270(.90)	268(.89)
		L	14(.05)	12(.04)	8(.03)	17(.06)
3.0	80%	H	20(.07)	27(.09)	36(.12)	25(.08)
		C	262(.87)	245(.82)	250(.83)	240(.80)
		L	18(.06)	28(.09)	14(.05)	35(.12)
	90%	H	13(.04)	14(.05)	20(.07)	11(.04)
		C	283(.94)	275(.92)	274(.91)	267(.89)
		L	4(.01)	11(.04)	6(.02)	22(.07)
4.0	80%	H	14(.05)	27(.09)	35(.12)	20(.07)
		C	278(.93)	245(.82)	248(.83)	239(.80)
		L	8(.03)	28(.09)	17(.06)	41(.14)
	90%	H	6(.02)	14(.05)	19(.06)	10(.03)
		C	286(.95)	275(.92)	274(.91)	262(.87)
		L	8(.03)	11(.04)	7(.02)	28(.09)

Table 4

Average Length of Confidence Intervals for ℓn P{D>t}
Exponential Sojourn Times
N=50, θ=0.5, ρ_1=1.0, ρ_2=10.0

Time	Confidence Level	BIN	MLE	JK	BT
0.5	80%	.216	.089	.155	.127
		(.002)	(.001)	(.004)	(.002)
	90%	.271	.114	.205	.162
		(.002)	(.001)	(.006)	(.003)
1.0	80%	.319	.177	.202	.181
		(.003)	(.001)	(.004)	(.002)
	90%	.401	.227	.268	.234
		(.003)	(.002)	(.005)	(.002)
1.5	80%	.416	.245	.282	.262
		(.003)	(.002)	(.005)	(.003)
	90%	.524	.340	.374	.340
		(.004)	(.003)	(.007)	(.004)
2.0	80%	.524	.353	.377	.352
		(.005)	(.003)	(.007)	(.004)
	90%	.628	.453	.500	.458
		(.006)	(.004)	(.009)	(.005)
3.0	80%	.750	.529	.585	.545
		(.008)	(.004)	(.011)	(.006)
	90%	.940	.579	.775	.706
		(.009)	(.005)	(.014)	(.008)
4.0	80%	1.05	.705	.799	.738
		(.015)	(.006)	(.015)	(.009)
	90%	1.31	.905	1.06	.959
		(.019)	(.007)	(.020)	(.011)

Table 5

Coverage Results for Confidence Intervals for ℓn P{D>t}
Nonexponential Sojourn Times
N=20, θ=0.5, ρ_1=2.0, ρ_2=10.0, ρ_3=2.0

Time	Confidence Level	Coverage	BIN	MLE (Exp)	JK	BT
0.5	80%	H	12(.04)	0(.0)	105(.35)	135(.45)
		C	278(.93)	53(.18)	183(.61)	162(.54)
		L	10(.03)	247(.82)	12(.04)	3(.00)
	90%	H	0(.00)	0(.00)	73(.24)	82(.37)
		C	290(.97)	109(.36)	220(.73)	217(.72)
		L	10(.03)	191(.64)	7(.02)	1(.0)
1.0%	80%	H	16(.05)	1(.0)	45(.15)	32(.11)
		C	254(.85)	221(.74)	240(.80)	241(.80)
		L	30(.10)	78(.26)	15(.05)	27(.09)
	90%	H	5(.02)	0(.0)	29(.10)	11(.04)
		C	285(.95)	266(.89)	263(.88)	271(.90)
		L	10(.03)	34(.11)	8(.02)	18(.06)
1.5	80%	H	21(.07)	8(.03)	40(.13)	23(.08)
		C	260(.87)	257(.86)	244(.81)	232(.77)
		L	19(.06)	35(.12)	16(.05)	45(.15)
	90%	H	5(.02)	2(.01)	24(.08)	7(.02)
		C	276(.92)	283(.94)	273(.91)	264(.88)
		L	19(.06)	15(.05)	3(.01)	29(.10)
2.0	80%	H	11(.04)	17(.06)	39(.13)	18(.06)
		C	270(.90)	260(.87)	246(.82)	232(.77)
		L	19(.06)	23(.08)	15(.05)	50(.17)
	90%	H	11(.04)	7(.02)	19(.06)	7(.02)
		C	284(.95)	284(.95)	277(.92)	263(.88)
		L	5(.02)	9(.03)	4(.01)	30(.10)
3.0	80%	H	13(.04)	31(.10)	37(.12)	17(.06)
		C	273(.91)	254(.85)	251(.84)	230(.77)
		L	14(.05)	15(.05)	12(.04)	53(.18)
	90%	H	5(.02)	17(.06)	19(.06)	6(.02)
		C	281(.94)	279(.93)	277(.92)	261(.87)
		L	14(.05)	4(.01)	4(.01)	33(.11)
4.0	80%	H	4(.01)	44(.15)	38(.13)	17(.06)
		C	279(.93)	245(.82)	251(.84)	229(.76)
		L	17(.06)	11(.03)	11(.04)	54(.18)
	90%	H	4(.01)	23(.08)	19(.06)	6(.02)
		C	296(.98)	273(.91)	277(.92)	260(.87)
		L	0(.00)	4(.01)	4(.01)	34(.11)

Table 6

Average Length of Confidence Intervals for $\ell n\ P\{D>t\}$
Nonexponential Sojourn Times
$N=20,\ \theta=0.5,\ \rho_1=2.0,\ \rho_2=10.0\ \rho_3=2.0$

Time	Confidence Level	BIN	MLE (Exp)	JK	BT
0.5	80%	.289	.144	.141	.130
		(.004)	(.002)	(.003)	(.003)
	90%	.359	.185	.183	.175
		(.005)	(.002)	(.004)	(.004)
1.0	80%	.485	.286	.269	.253
		(.007)	(.003)	(.006)	(.004)
	90%	.604	.367	.351	.331
		(.008)	(.004)	(.008)	(.006)
1.5	80%	.681	.428	.451	.419
		(.010)	(.005)	(.010)	(.007)
	90%	.850	.550	.588	.552
		(.013)	(.007)	(.014)	(.010)
2.0	80%	.884	.571	.643	.594
		(.013)	(.007)	(.015)	(.017)
	90%	1.10	.733	.838	.786
		(.017)	(.009)	(.019)	(.015)
3.0	80%	1.46	.855	1.03	.953
		(.034)	(.010)	(.024)	(.018)
	90%	1.81	1.10	1.35	1.26
		(.043)	(.013)	(.032)	(.025)
4.0	80%	2.34	1.14	1.43	1.31
		(.079)	(.013)	(.034)	(.025)
	90%	2.84	1.46	1.86	1.74
		(.084)	(.017)	(.044)	(.035)

Table 7

Coverage Results for Confidence Intervals for $\ell n\ P\{D>t\}$
Nonexponential Sojourn Times
$N=50,\ \theta=0.5,\ \rho_1=2.0,\ \rho_2=10.0,\ \rho_3=2.0$

Time	Confidence Level	Coverage	BIN	MLE (Exp)	JK	BT
0.5	80%	H	18(.06)	0(.0)	130(.43)	149(.50)
		C	261(.87)	3(.01)	165(.55)	150(.50)
		L	21(.07)	297(.99)	5(.02)	1(.0)
	90%	H	10(.03)	0(.00)	99(.33)	106(.35)
		C	283(.94)	6(.02)	200(.67)	194(.65)
		L	7(.02)	294(.98)	1(.00)	0(.0)
1.0	80%	H	22(.07)	2(.01)	42(.14)	39(.13)
		C	260(.87)	145(.48)	242(.81)	240(.80)
		L	18(.06)	153(.51)	16(.05)	21(.07)
	90%	H	11(.04)	0(.00)	28(.09)	21(.07)
		C	280(.93)	213(.71)	263(.88)	265(.88)
		L	9(.03)	87(.29)	9(.03)	14(.05)
1.5	80%	H	27(.09)	6(.02)	31(.10)	25(.08)
		C	254(.85)	242(.81)	247(.82)	243(.81)
		L	19(.06)	52(.17)	22(.07)	32(.11)
	90%	H	11(.04)	4(.01)	21(.07)	19(.06)
		C	276(.92)	276(.93)	271(.90)	262(.87)
		L	13(.04)	20(.07)	8(.02)	19(.06)
2.0	80%	H	25(.08)	13(.04)	29(.10)	24(.08)
		C	253(.84)	264(.88)	250(.83)	239(.80)
		L	22(.07)	23(.08)	21(.07)	37(.12)
	90%	H	16(.05)	7(.02)	20(.07)	12(.04)
		C	274(.91)	282(.94)	269(.90)	264(.88)
		L	10(.03)	11(.04)	11(.04)	24(.08)
3.0	80%	H	25(.08)	34(.11)	31(.10)	21(.07)
		C	247(.82)	253(.84)	248(.83)	241(.80)
		L	28(.09)	13(.04)	21(.07)	38(.13)
	90%	H	17(.06)	20(.07)	18(.06)	11(.04)
		C	273(.91)	275(.92)	271(.90)	266(.89)
		L	10(.03)	5(.02)	11(.04)	23(.08)
4.0	80%	H	11(.04)	47(.16)	31(.10)	20(.07)
		C	266(.89)	246(.82)	244(.81)	238(.79)
		L	23(.08)	7(.02)	25(.08)	42(.14)
	90%	H	6(.02)	29(.10)	19(.06)	10(.03)
		C	285(.95)	267(.89)	270(.90)	264(.88)
		L	9(.03)	4(.01)	11(.04)	26(.09)

Table 8

Average Length of Confidence Intervals for ℓn P{D>t}
Nonexponential Sojourn Times
N=50, θ=0.5, ρ_1=2.0, ρ_2=10.0, ρ_3=2.0

Time	Confidence Level	BIN	MLE	JK	BT
0.5	80%	.169	.089	.091	.081
		(.002)	(.001)	(.002)	(.001)
	90%	.211	.114	.121	.104
		(.002)	(.001)	(.003)	(.002)
1.0	80%	.284	.176	.167	.153
		(.002)	(.001)	(.003)	(.002)
	90%	.357	.227	.221	.199
		(.003)	(.002)	(.004)	(.002)
1.5	80%	.356	.264	.269	.250
		(.003)	(.002)	(.005)	(.003)
	90%	.498	.339	.357	.324
		(.004)	(.002)	(.006)	(.004)
2.0	80%	.514	.352	.378	.351
		(.005)	(.003)	(.007)	(.004)
	90%	.646	.452	.501	.456
		(.006)	(.003)	(.010)	(.005)
3.0	80%	.784	.527	.600	.558
		(.009)	(.004)	(.011)	(.006)
	90%	.983	.677	.795	.726
		(.011)	(.005)	(.014)	(.009)
4.0	80%	1.16	.703	.824	.765
		(.019)	(.063)	(.015)	(.009)
	90%	1.45	.903	1.09	.998
		(.024)	(.064)	(.020)	(.012)

DISTRIBUTION LIST

	NO. OF COPIES
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Office of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analyses 4401 Ford Ave. Alexandria, VA 22311	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Operations Research Center, Rm E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J. F. Shapiro Cambridge, MA 02139	1
Koh Peng Kong OA Branch, DSO Ministry of Defense Blk 29 Middlesex Road SINGAPORE 1024	1
Arthur P. Hurter, Jr. Professor and Chairman Dept of Industrial Engineering and Management Sciences Northwestern University Evanston, IL 60201-9990	1
Institute for Defense Analysis 1800 North Beauregard Alexandria, VA 22311	1
Chief of Naval Research 800 N. Quincy St. Arlington, VA 22217-5000	1

END
FILMED
MARCH, 1988
DTIC